

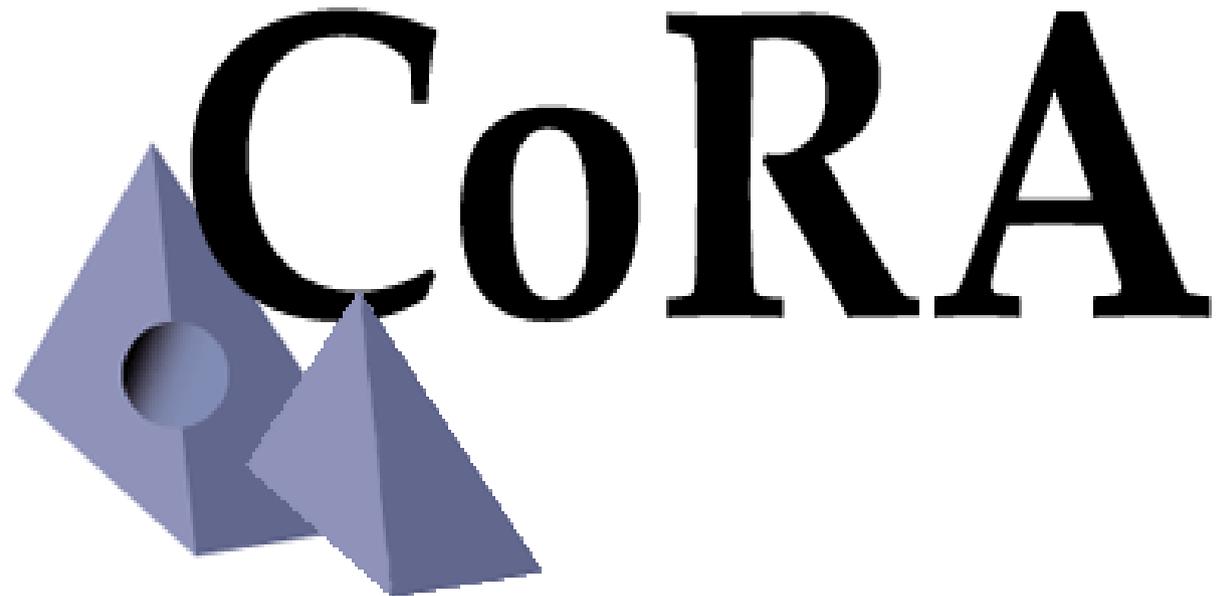
Solar Flare Prediction Using Discriminant Analysis



Jeff Tessein, KD Leka, Graham Barnes
NorthWest Research Associates Inc.
Colorado Research Associates Division

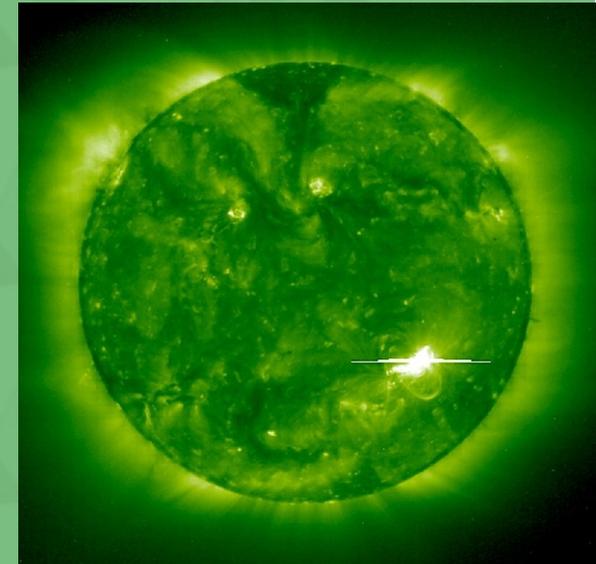
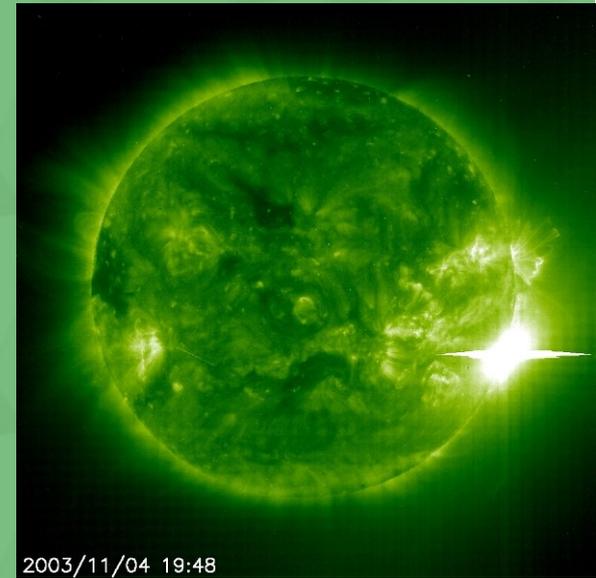
NWRA

Since 1984



Motivation for this work

- There is no clear indication of when a solar flare is going to occur
- Predictions are needed to minimize the harm inflicted by solar flares
 - GPS navigation rendered useless, blackouts, harm to astronauts, and possible damage to satellites
- Ability to predict “all clear”
 - 5% or lower chance of a major event
- The diverse need for flare predictions calls for different warning times
 - Planes flying polar routes might need more prep time in order to change their flight path than the power company needs to shut down vulnerable areas of the power grid



TRACE

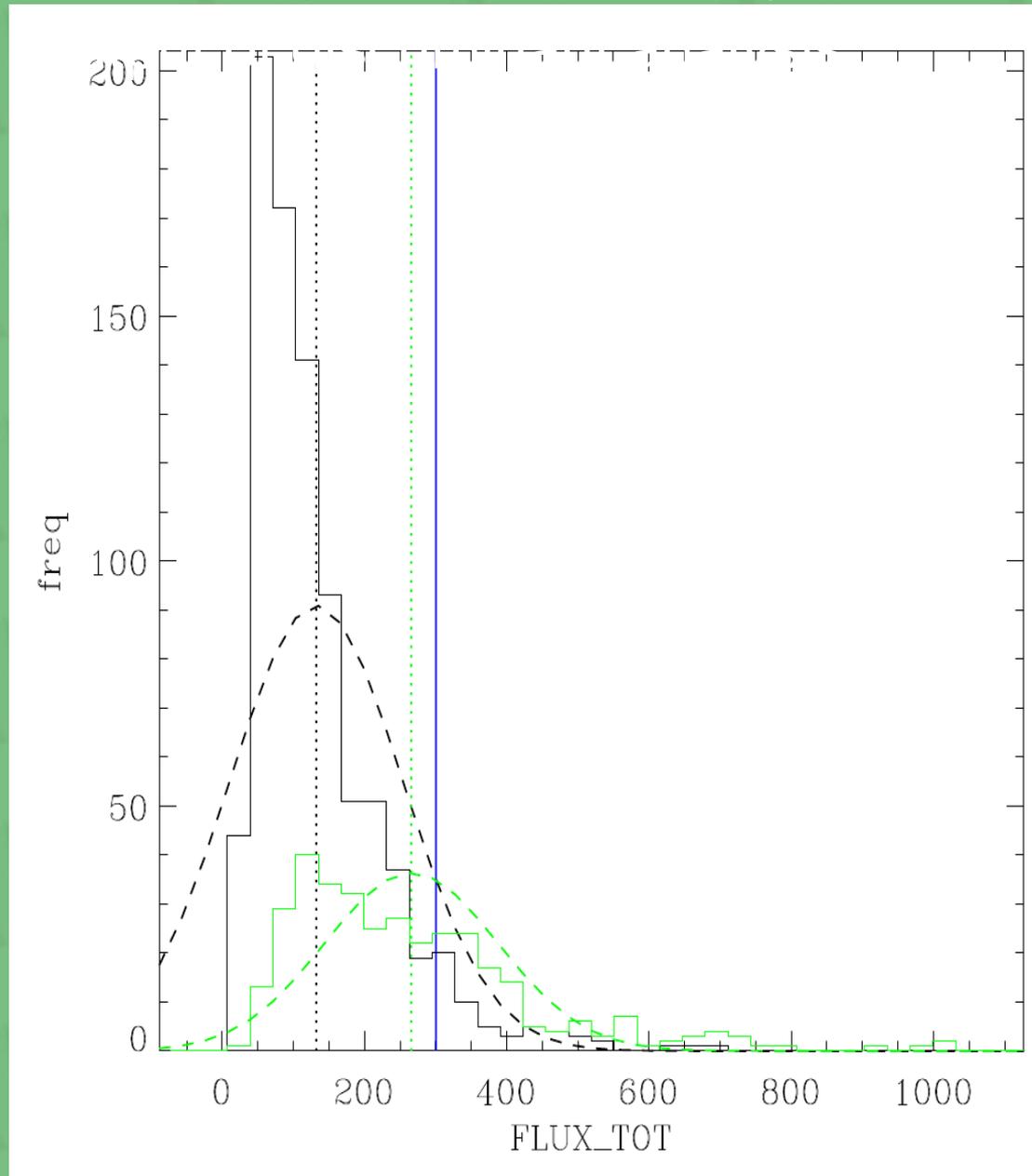
Solar Flare Prediction

- Lots of methods out there but none are absolutely effective, and we don't know exactly how effective they are
- The discriminant analysis can sort through these methods and compare them
- Goal is to find the most powerful predictors of solar flares
- We expect that this will be a group of parameters rather than just one
 - We have four that perform the best (both individually and grouped together): total flux, current helicity, vertical current, and the standard deviation of neutral line shear
- Look at different time intervals – my summer project
 - Nominally a flaring event is 24 hours
 - Analyze results of the analysis when a flaring interval is shorter

Discriminant Analysis

- Statistical technique that classifies cases into one of two groups (flaring and non-flaring)
 - Histograms represent populations
 - Blue line is the discriminant boundary
- Can consider many variables simultaneously
- Determine how flaring and non-flaring intervals are statistically separated
- Climatology skill score is a measure of improvement over predicting the same thing all of the time

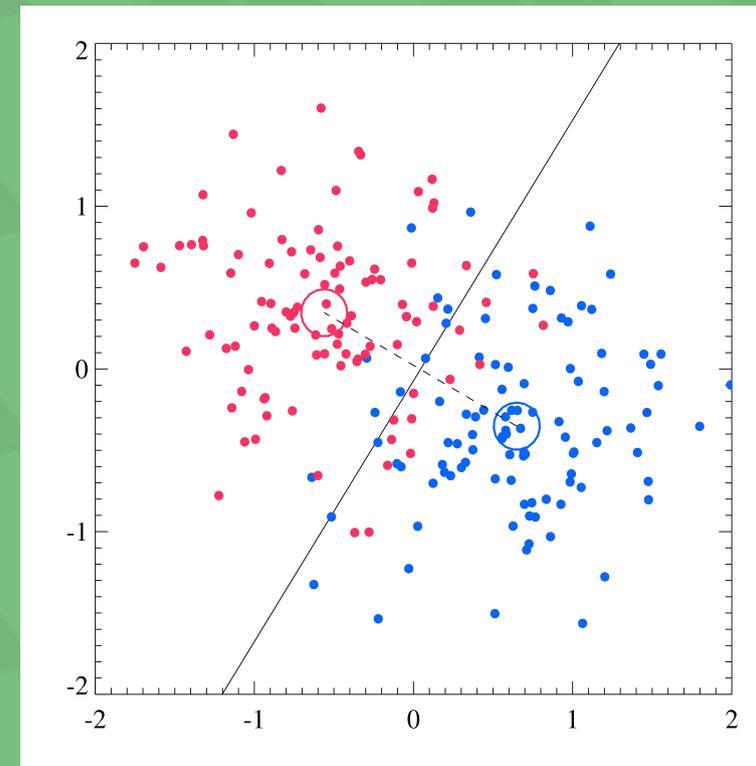
Output of discriminant analysis



Discriminant Analysis

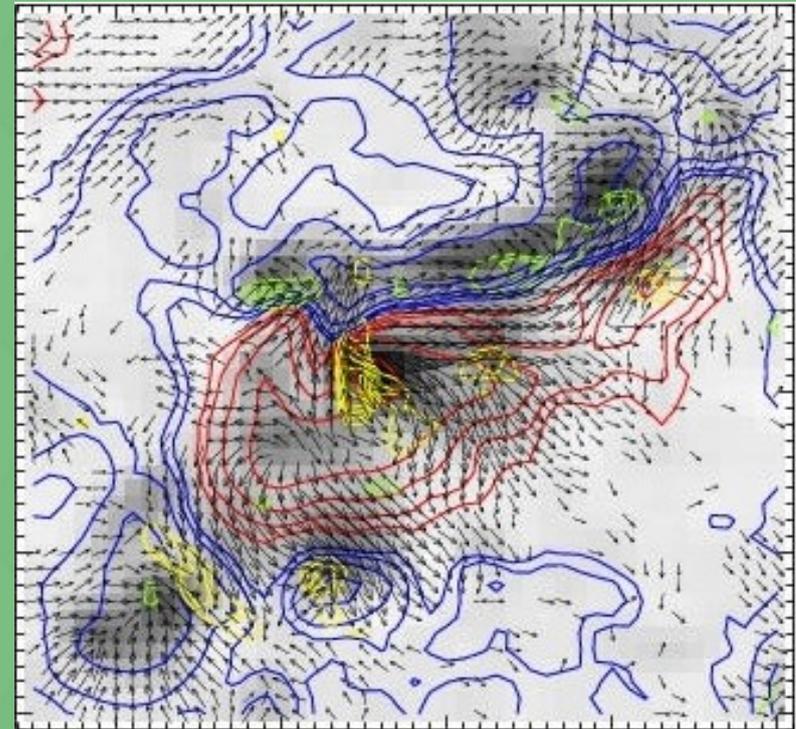
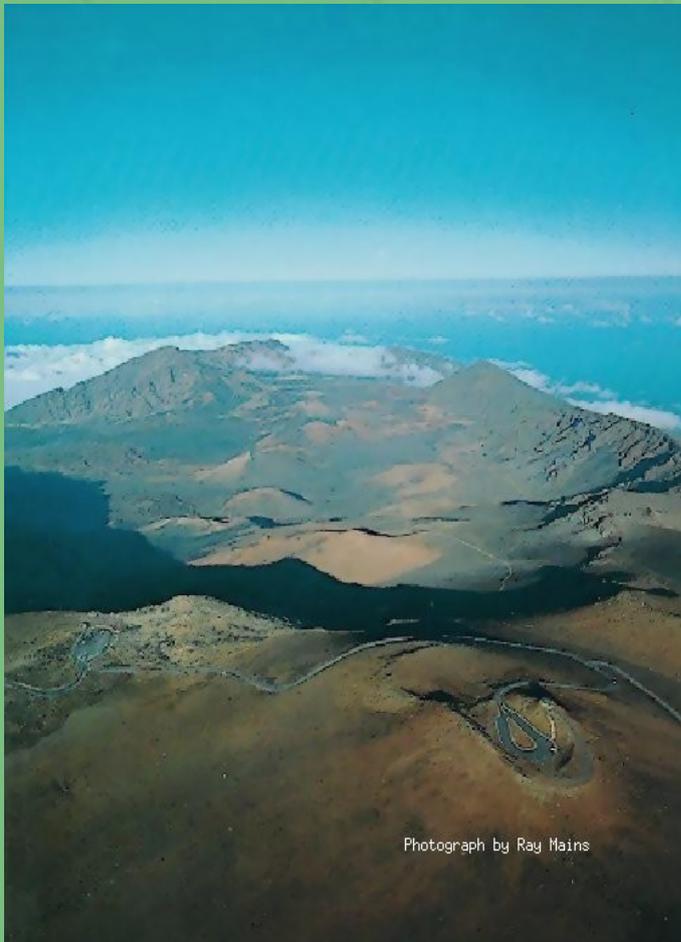
- Classification table shows how well the analysis has performed
- Diagonal elements are correct predictions and off diagonals are incorrect predictions
- Fake data set depicts two different populations
 - Diagonal line is discriminant boundary or “50% line”
 - Circles are means
 - Distance between them is Mahalanobis distance – indicates amount of statistical difference between the samples
- In solar flare context, a point on the discriminant boundary statistically has a 50% chance of flaring
- When running the DA, there are just a few parameters affecting the analysis, and as we add more and more variables, the results will not improve past a certain point.

		Predicted	
		F	NF
Observed	F	117	229
	NF	49	817



Data

- Magnetograms from the IVM (Imaging Vector Magnetograph) at Mees Solar Observatory in Hawai'i
 - The IVM can observe magnetic field vectors in the photosphere
 - Magnetograms are taken every day



Vector magnetogram



Haleakala

Data

- Magnetograms are used to derive other parameters used in the analysis e.g. Total flux
- Some parameters are not used because they are repetitive
- Flaring intervals are determined by:
 - Flare occurring within a certain time after the magnetogram
 - Flare exceeds a predetermined minimum soft X-ray flux

$$\Phi_{tot} = \Sigma |B_z| dA$$

$$h_c = CB_z \left(\frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y} \right)$$

$$J_z = \frac{\left(\frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y} \right)}{\mu_0}$$

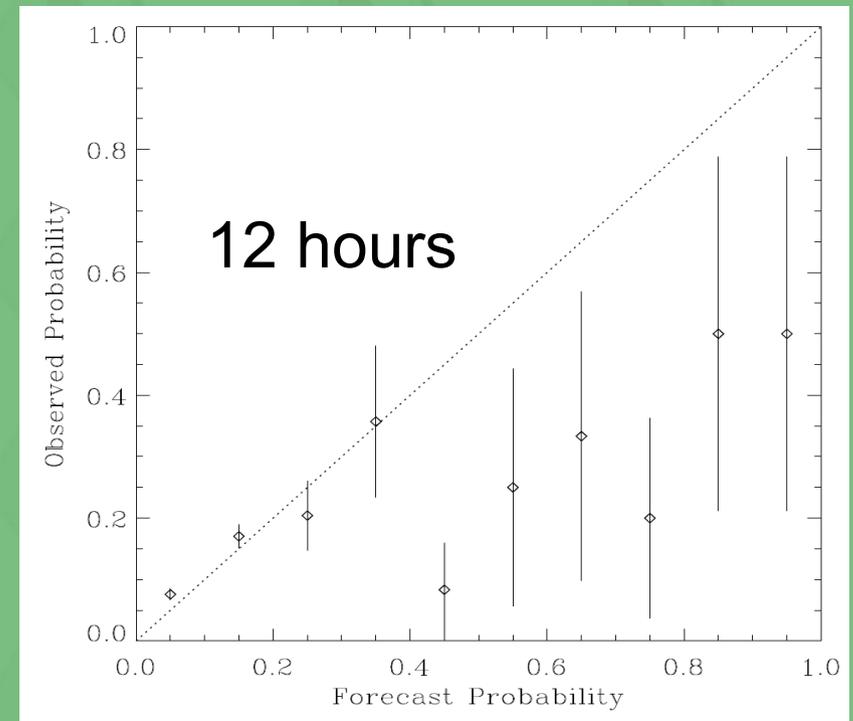
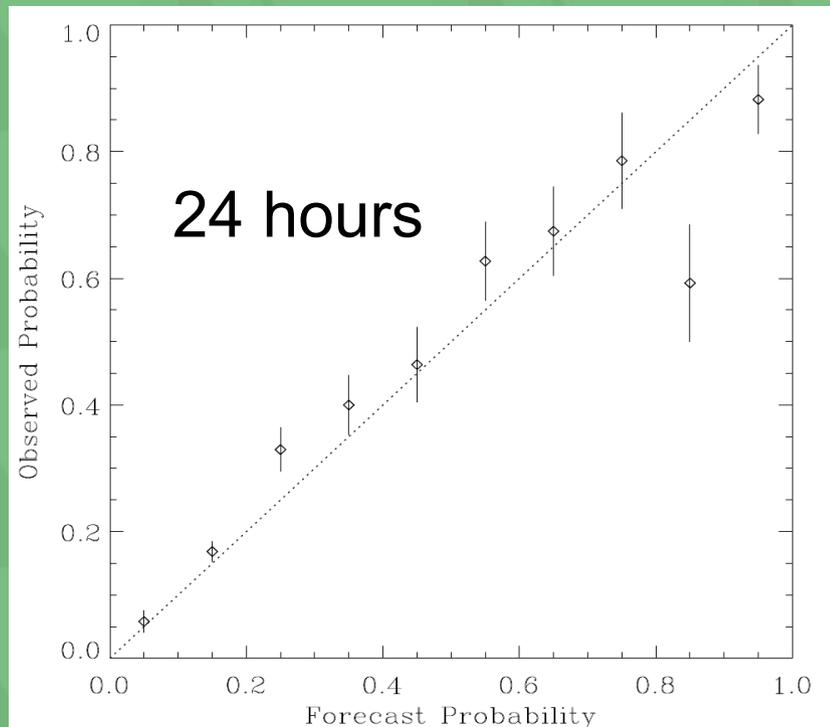
$$\Psi_{NL} = \arccos \left(\frac{\mathbf{B}_{NL}^p \cdot \mathbf{B}_{NL}^0}{|\mathbf{B}_{NL}^p| |\mathbf{B}_{NL}^0|} \right)$$

Sample data:

10039.000	20020725.	16.916666	
5			
20020725.	18.983334	2.5000000e-06	0.0000000
20020725.	21.116667	2.8000000e-06	1.0000000
20020725.	21.916666	3.1000000e-06	0.0000000
20020725.	22.900000	6.7000000e-06	0.0000000
20020726.	0.066666670	4.9000000e-05	0.0000000

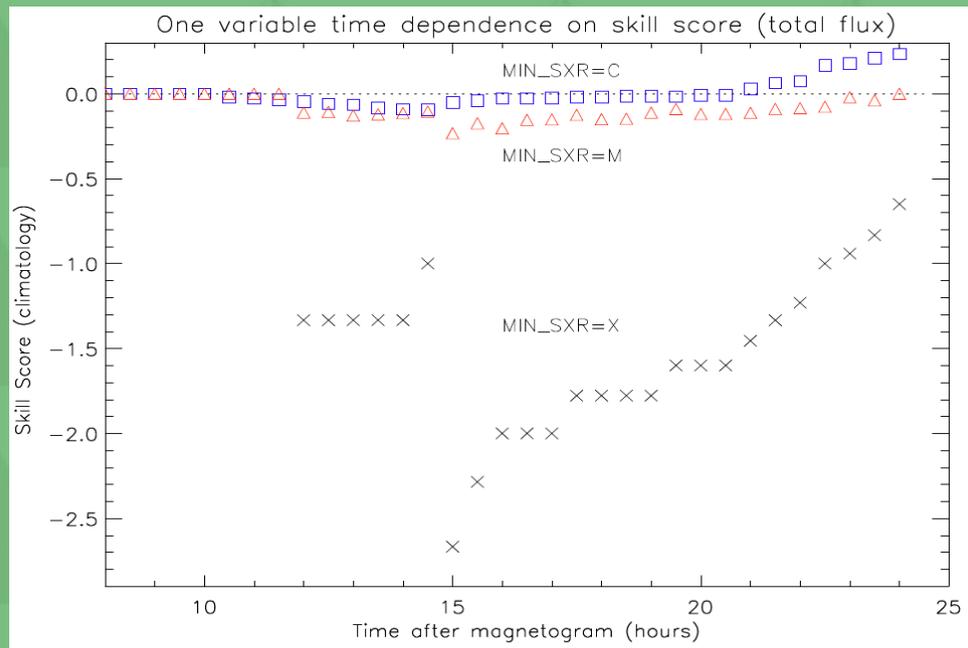
ΔT analysis

- Analysis based on variable time intervals
- This analysis will allow us to see how the predictive power for magnetograms changes as the time of measurement becomes closer to the flare time
- 12 hour analysis might be less reliable due to small population



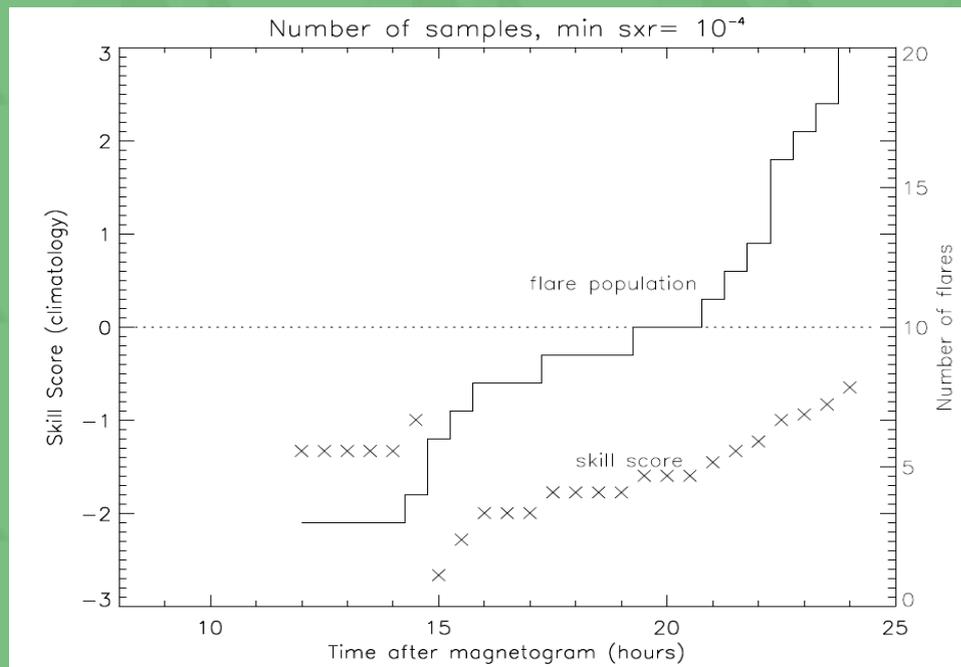
These are reliability plots. The dashed line is an ideal prediction based on the probability that an active region is going to flare.

ΔT analysis



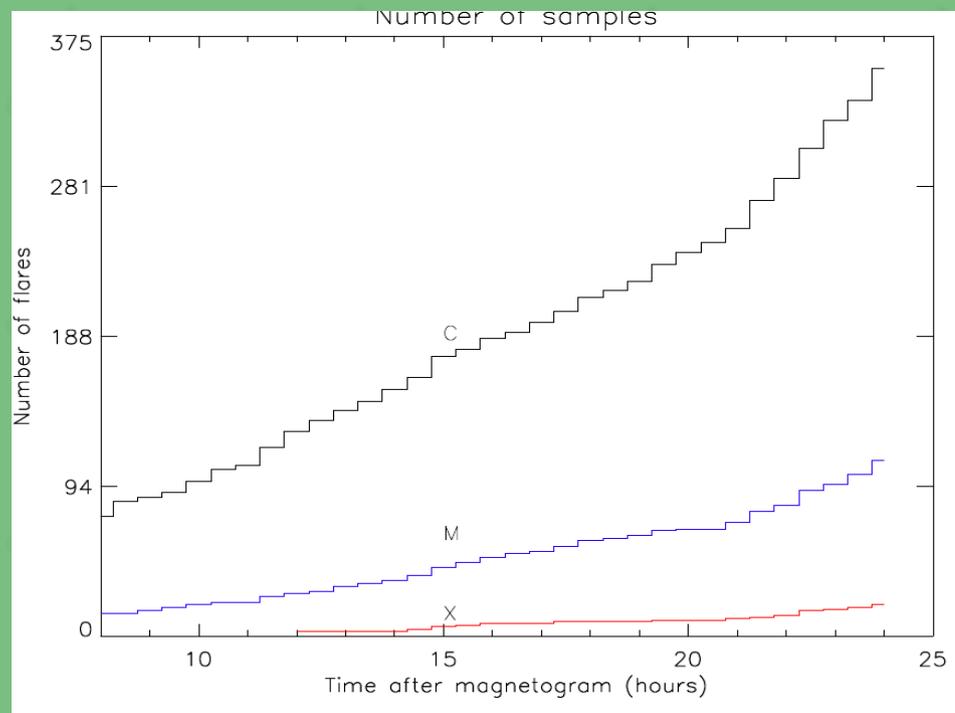
- To the left is skill score plotted as a function of ΔT
- Skill score appears to increase toward 24 hours
- The dashed line at skill score=0 is plotted for clarity

- The plot to the right indicates how small populations negatively affect the analysis
- Minimum SXR flux of 10^{-4} (X flare) is an extreme example

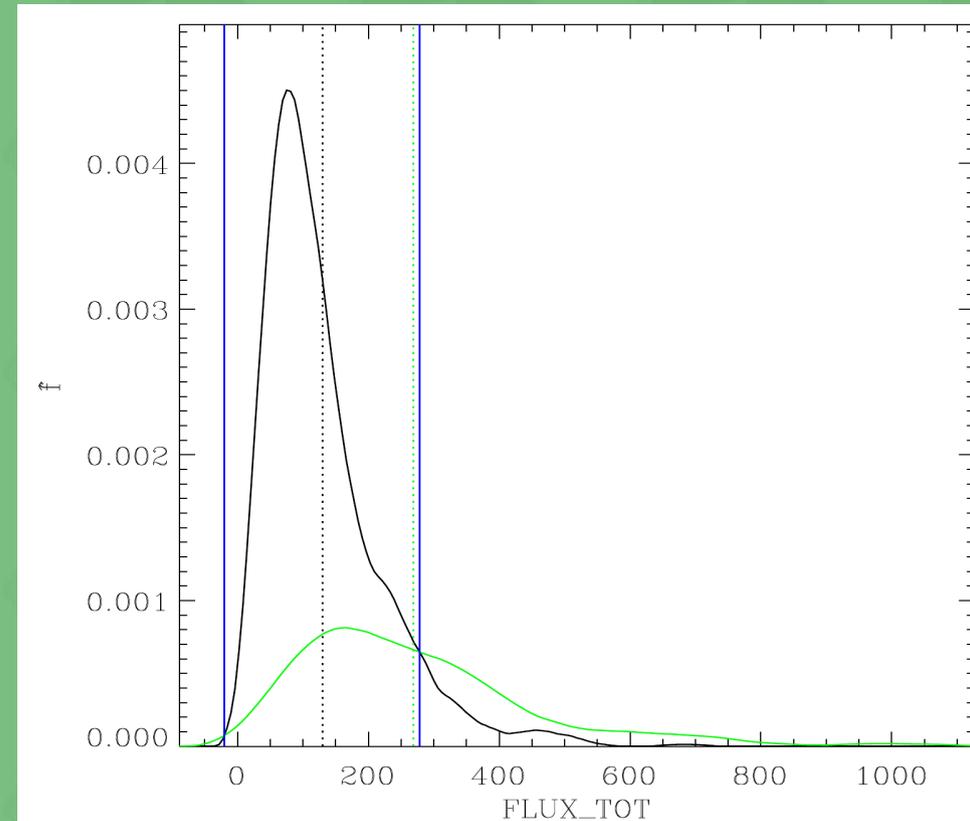
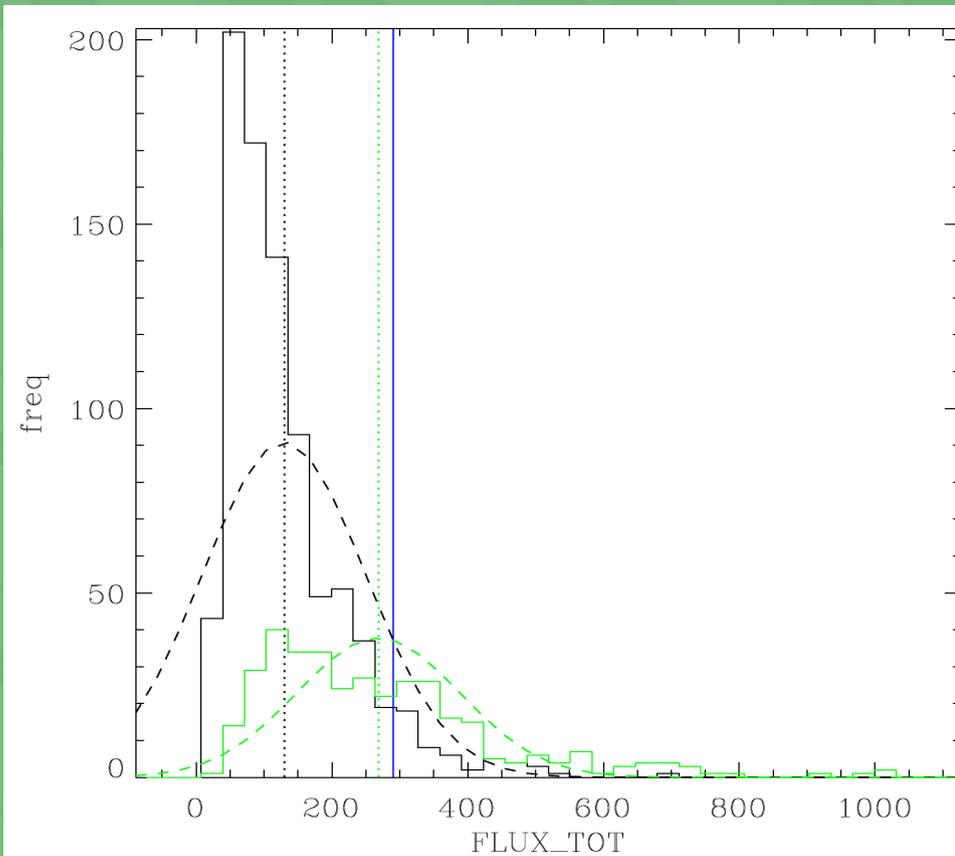


ΔT analysis

- The smallest time intervals we can look at depend on sample size
- A higher SXR flux will have a smaller sample size
- At extremely small time intervals there is a very lopsided population
 - This weakens any statistical analysis
- For this reason it's difficult to know what results the analysis will find at small time intervals

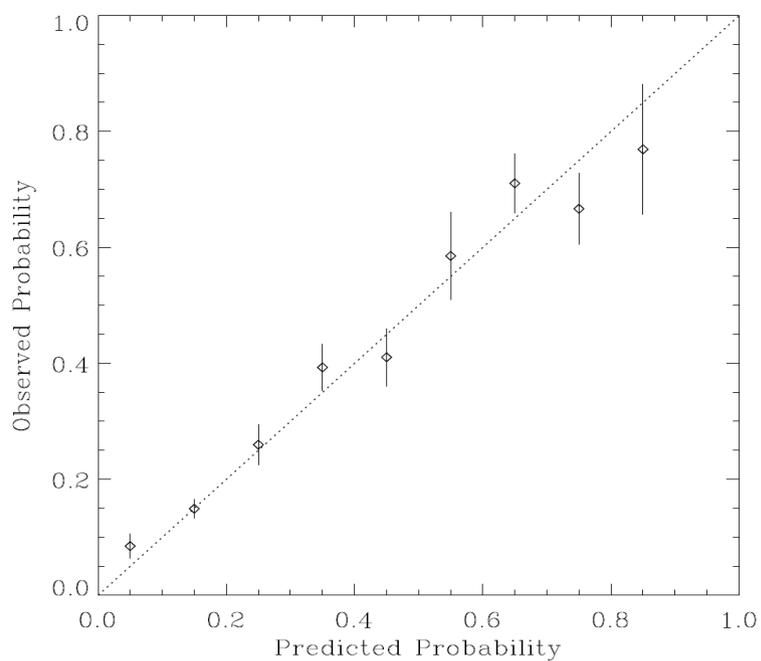


Nonparametric Discriminant Analysis

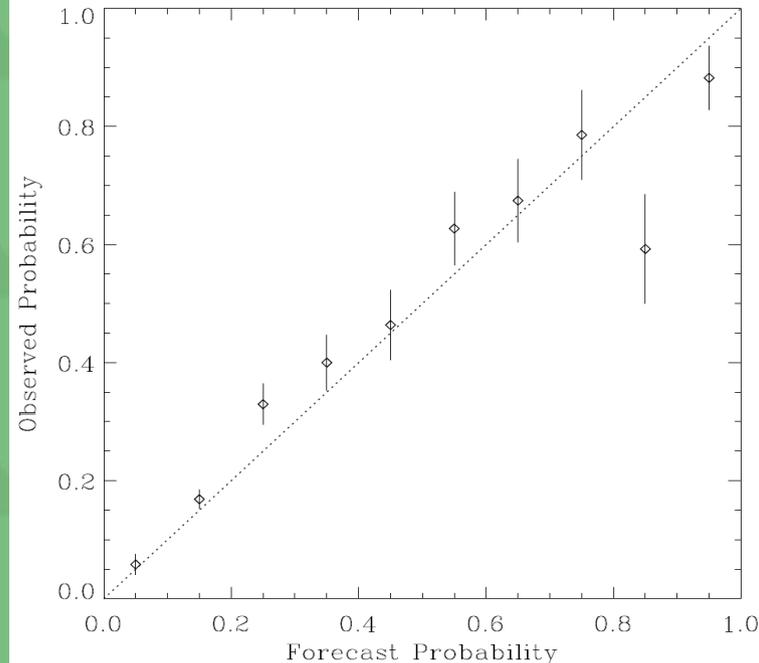


- An alternate option to the discriminant analysis
- The linear discriminant analysis uses a Gaussian while the NDA uses a different smoothing function. In the case of flares they will have different tails
- NDA shown on the right
- Needs large sample sizes to reach full potential

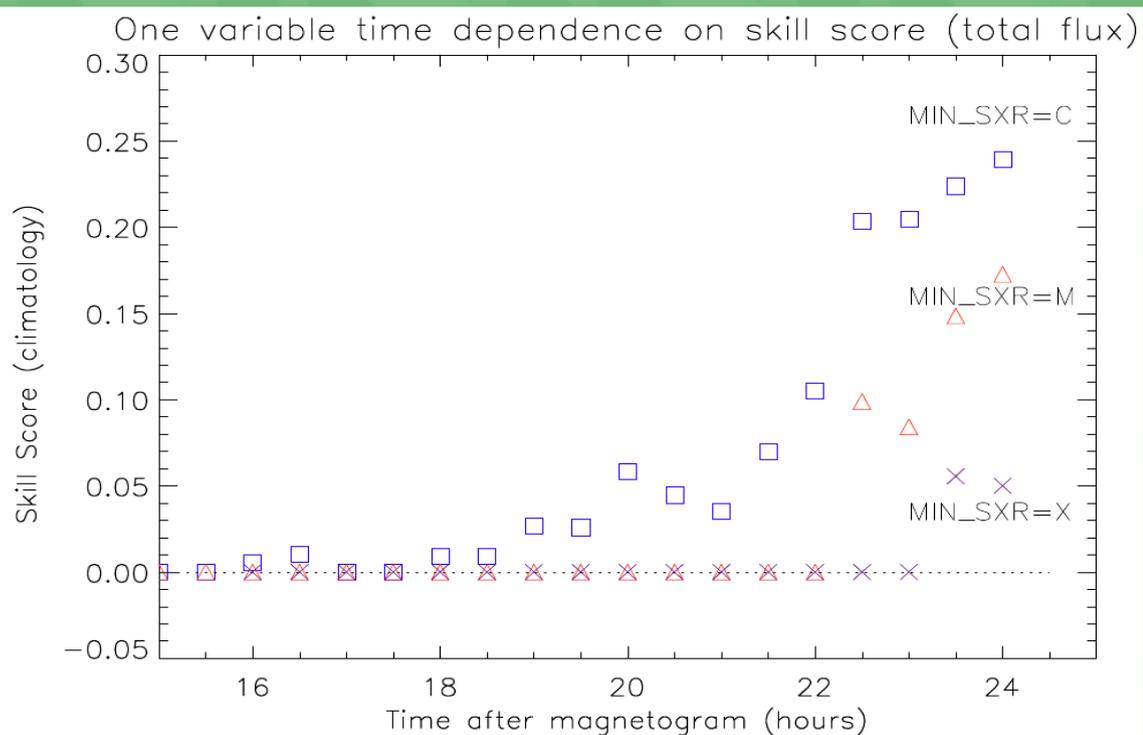
Nonparametric Discriminant Analysis



Nonparametric DA is top left – it performs slightly better than the linear DA (top right).

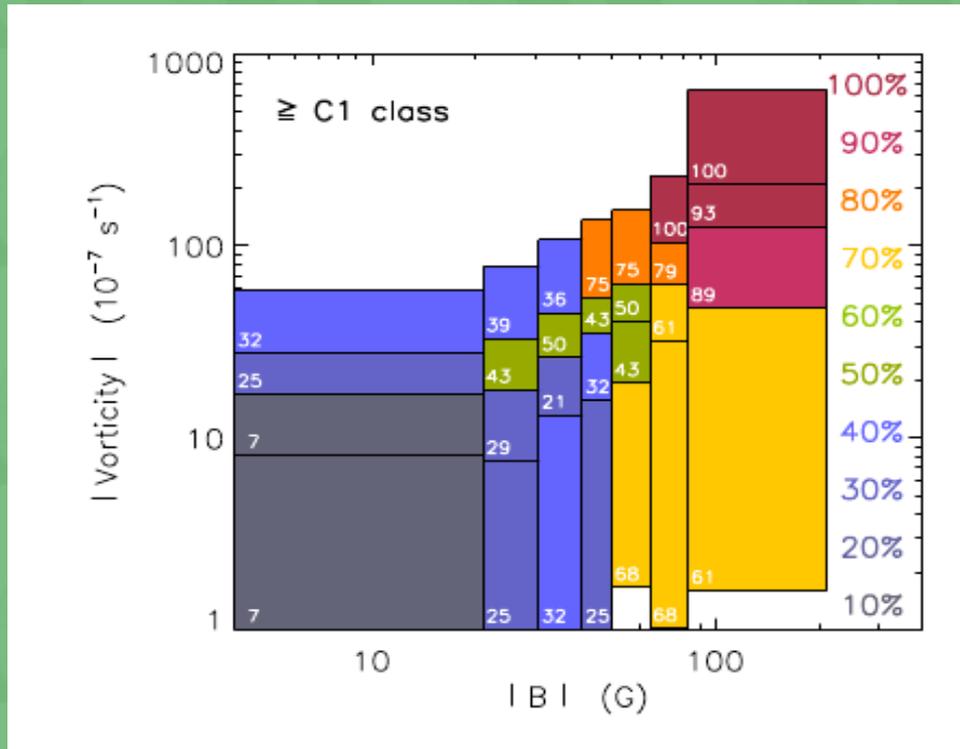
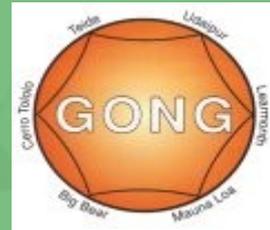


NDA ΔT results are shown to the right. These are C-and-above flares, which includes as much data as possible. By looking at the difference between 18 and 24 hours, it would appear that a forecast at small ΔT is not possible.



Discriminant Analysis using GONG data

Global Oscillation Network Group; A network of stations all over the world with a constant view of the sun providing helioseismology data. The data used in this analysis include subsurface plasma flow, subsurface vorticity, magnetic field, and flaring history.



Left: This data has been used to show that subsurface vorticity can be useful in flare prediction.

The results from this analysis are impressive.

- BMAX is making an overwhelming contribution to the analysis
- Actual helioseismology data doesn't seem to be doing much
- Skill score using top two parameters is almost as high as when all twelve are used
- Some of the GONG data does worse than latitude and longitude

classification table:

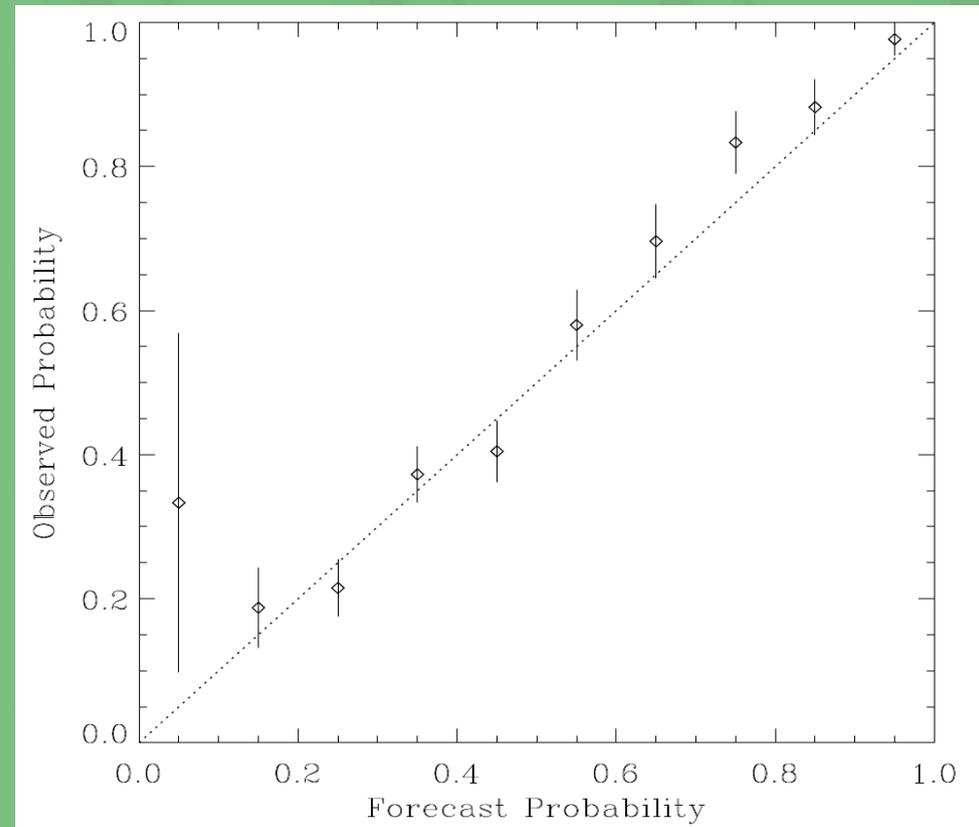
270	138
82	294

rate of correct classification:
0.719388

Heidke skill score (climatology):
0.460784

Heidke skill score (random):
0.441103

```
discriminant function coefficients
 1 BMAX          1.03790
 2 OMEGA_S_DEEP_ 0.421854
 3 OMEGA_S_      -0.336330
 4 OMEGA_X_DEEP_ -0.247244
 5 OMEGA_Y_DEEP_ -0.232098
 6 RADIUS        0.217511
 7 OMEGA_X_      -0.125590
 8 LAT_          -0.101354
 9 OMEGA_Y       0.0864090
10 OMEGA_Z_      -0.0332914
11 LONG          -0.0315747
12 OMEGA_Z_DEEP_ 0.0135476
```



Summary

- It's clear that small number statistics are a big issue.
- These results show that skill scores can improve with more data points.
- Beyond that, the samples aren't large enough to say with confidence what impact small ΔT has on forecasting.
- A larger database may be introduced in the future in attempt to rectify this.
- NDA ΔT results are shown to the right. These are C-and-above flares, which includes as much data as possible. By looking at the difference between 18 and 24 hours, it would appear that a forecast at small ΔT is not possible.

SOHO/MDI image of the sun
on July 30, 2008

